A geoadditive distributional regression analysis of the local relationship of land prices and land rents in Germany

Henning Schaak, Oliver Musshoff

FORLand-Working Paper 20 (2020)

Agricultural Land Markets – Efficiency and Regulation

Published by
DFG Research Unit 2569 FORLand, Humboldt-Universität zu Berlin
Unter den Linden 6, D-10099 Berlin
https://www.forland.hu-berlin.de

Tel +49 (30) 2093 46845, Email gabriele.wuerth@agrar.hu-berlin.de

A geoadditive distributional regression analysis of the local relationship of land prices and land rents in Germany

Henning Schaak *, Oliver Musshoff **

January 2020

Abstract

The analysis of sales prices and rents for agricultural land are a classical research topic in agricultural economics. Due to increased dynamic on agricultural land markets, their relationship has gained increased interest recently. The present study contributes to the literature by studying the district level heterogeneity of the rent-price-ratio (RPR) of agricultural land. This was achieved by modelling the full conditional distribution of the RPR, using a generalised additive model for location, shape and scale (GAMLSS). In order to choose an adequate model specification; a variable selection procedure is applied. The analysis utilised data from the German federal state Lower Saxony, containing all observable sales and rent price data. Shares of different field crops, livestock densities, shares of different farm types, and the concentration of land were found to influence the distribution of the RPR. Furthermore, differences in the distribution between arable land and grassland were found. By explicitly modelling and visually presenting spatial effects, additional insights into the spatial variation of the profitability of investments in farmland in Germany are provided. Thereby, conclusions regarding efficiency of land markets are possible.

Keywords: Agricultural land markets; price diffusion; spatial dependence; border effect

JEL codes: C21, C46, Q15

Acknowledgements

The authors gratefully acknowledge financial support from the German Research Foundation (DFG) through Research Unit 2569 “Agricultural Land Markets – Efficiency and Regulation”. They also thank the members of the FDZ in Kiel, especially Dr. Steffi Dierks for their outstanding assistance and OGA Lower Saxony, especially Peter Ache, for proving data access.

* **Henning Schaak**, henning.schaak@agr.uni-goettingen.de

** **Oliver Musshoff**, oliver.musshoff@agr.uni-goettingen.de

Both authors are at University of Goettingen, Department for Agricultural Economics and Rural Development, Platz der Göttinger Sieben 5, 37073 Göttingen, Germany.

1 Introduction

Land is the central factor of agricultural production. The developments of prices for agricultural land and their determinants are therefore an important research topic in agricultural economics. Sharp price increases in Germany (Statistisches Bundesamt 2018), as well as other European countries (Eurostat 2019) motivated intensive research activity. Latruffe and Mouël (2009) reported that land prices were positively affected by agricultural support policy instruments. Likewise, Hennig et al. (2014) found a positive effect of payment entitlements on land rental prices. The effect of biogas subsidies on rental rates was studied by Habermann and Breustedt (2011) and Hennig and Latacz-Lohmann (2017). Feichtinger and Salhofer (2013) provided a meta-analysis on the impact of subsidies on agricultural land prices. Regarding the spatial dynamics of land prices, Yang et al. (2017) found regional convergence clusters in Germany. In the follow of the financial crisis 2008, the hypothesis, that additional demand from non-agricultural investors has accelerated price increases, has drawn attention (Tietz et al. 2013; Hüttel et al. 2016; Plogmann et al. 2018). However, as land can be bought and rented, ideally research would have to consider both sales prices and rental rates for land, as well as their dependencies.

While there are multiple ways to approach the analysis of farmland values, it commonly is rooted in the theory of valuing financial assets which is dependent on income capitalisation, or the net present value (NPV) model (Burt 1986). In this context, theoretical land values were often derived by using cash rents as a proxy for returns from agricultural activities. In an efficient market, the sales price should equal the capitalised returns and therefore further only depend on interest rate. These theoretical farmland values can be compared with observed values. Alternatively, the ratio between observed rental and sales prices (rent-price ratio; RPR) can serve as an indicator of the profitability of an investment in land. Likewise, under the assumption of a static economic environment, the ratio can be interpreted as the salvage factor of an investment in land. Using a variance decomposition approach (Campbell and Shiller 1988), the RPR was recently studied by Plogmann et al. (2018) on the federal states level in the German land market. The authors found substantial variation of the RPR between federal states, which remains unexplained. While these results potentially could be explained by regional differences in the farming structure and the natural conditions. However these dimensions can also vary substantially within a given region. Therefore, the question arises, how heterogeneous the cross-sectional RPR is on finer spatial levels and whether it can be explained by the local farm structure.

The relationship of agricultural land prices and rental rates has been extensively studied by economic research (see e.g. Hyder and Maunder 1974; Traill 1979; Phipps 1984; Alston 1986; Burt 1986; Falk 1991). Generally, it is assumed that cash rents should vary in together with farmland values, with a strong positive relationship in their respective trends (Gutierrez et al. 2007). Ibendahl and Griffin (2013) found that rents lag behind changes in land prices when they are increasing, but not when they are decreasing. On the other hand, Saguatti et al. (2014) found that the long-run elasticity of cropland values with respect to net cash rents was close to unity. This can be interpreted as evidence for the validity of the NPV assumption. However, the literature also suggests there are conflicting results. Although farmland price and rental rate movements are highly correlated, price movements are not always in accordance with the expected relationship (Falk 1991; Hallam et al. 1992; Clark et al. 1993). Therefore, also the real options approach has been applied, in order to account for uncertainty in future growth and capital gains (Turvey 2002).

An important related issue is the (cross-sectional) heterogeneity on the land markets, as land remains an important cost factor in agricultural production. Thus the understanding of the land market is important for the proper understanding of other production-related developments. Nevertheless, heterogeneity of agricultural land sales prices and farmland rental rates is an issue that has rarely been considered in the literature (Mishra and Moss 2013; März et al. 2016). Even if the NPV approach holds, farmland values could obviously vary with different natural conditions. Still, at a given interest rate, the RPR should be identical between regions. Taking the work of Plogmann et al. (2018), who found that this is not the case in Germany, as a starting point, the present paper revisits the relationship between rental rates and sales prices on basis of the RPR. In contrast to Plogmann et al. (2018), this paper focusses on the cross-sectional distribution of the RPR on the district level and additionally addresses its heterogeneity within the individual districts. The paper is the first to explicitly model all parameters of the district level distribution of the RPR. It uses a unique dataset, combining data from the German agricultural census and data collected by the expert committees for land evaluation Lower Saxony (Oberer Gutachterausschuss für Grundstückswerte Niedersachsen, OGA Lower Saxony). The study area is well-suited for the research topic, as local farming structures in Lower Saxony are heterogeneous, with areas of intensive dairy, livestock and crop production. Other structural parameters, like the average farm size varied considerable at the local level (cf. NMELV 2017; Destatis n.d.).

In order to model the parameters of the RPR's distribution, the paper relies on the GAMLSS-framework (Rigby and Stasinopoulos 2005). In this framework, not only the mean, but also the higher moments of a distribution could be modelled by generalized additive models (GAMs; Hastie and Tibshirani 1990). Thus, the response distribution is completely characterized by one joint model (Umlauf and Kneib 2018). In this context, spatial dependencies and effects on the district level can be modelled by structured as well as unstructured spatial effects (Fahrmeir and Kneib 2011). Modelling the mean and the scale parameter of the RPR's distribution allowed identifying factors which influence the average profitability of an investment in land, as well as its heterogeneity.

The remainder of this paper is structured as follows: The second section outlines the methodological basis of the paper. In the third section, the used datasets and their preparation are described, followed by a motivation of the applied variable selection procedure. The results of the analysis are presented and discussed in section four. The paper ends with conclusions (section 5).

2 Methodology

In order to model the distribution of the RPR, all parameters of its conditional distribution are considered. In context of regression methods, this can be achieved by using a generalisation of the GAM-framework, referred to as "generalized additive models for location shape and scale" (GAMLSS; Rigby and Stasinopoulos 2005). The GAMLSS-framework makes way for more flexibility as more traditional regression frameworks. This is achieved by (a) not only modelling the mean (or location) parameter of the dependent variable's conditional distribution but also other parameters (e.g. the variance) and (b) considering non-exponential distributions. The aim of this section is not to give a comprehensive presentation of the framework, but rather to outline the overall concept and the specification used in the present study. For general discussions of the GAMLSS-framework, the reader is referred to the canonical references (e.g. Rigby and Stasinopoulos 2005; Stasinopoulos and Rigby 2007; Stasinopoulos et al. 2017).

Generally, within the GAMLSS-framework, the p parameters $\boldsymbol{\theta}_k = (\theta_1, \theta_2, \ldots, \theta_p)$ of the dependent variable Y's distribution are individually modelled by a GAM. As mentioned earlier, the distribution of Y is not limited to the exponential family and can be chosen from a more general family (cf. Rigby et al. 2019). Distributions from this family have up to 4 parameters which can be modelled. Depending on the specific distribution, the parameters represent the distributions location (e.g. the mean), scale (e.g. the variance) and shape (skewness and kurtosis). It is assumed that for $i = 1,2,\ldots,n$, independent observations Y_i have the probability density function $f_Y(y_i|\boldsymbol{\theta}^i)$, where $\boldsymbol{\theta}^i = {\boldsymbol{\theta}_k}^T$. The original formulation of the GAMLSS is described in the following (Rigby and Stasinopoulos 2005).

Using a known monotonic link function $g_k(\cdot)$ (e.g. a log-link function) to relate $\boldsymbol{\theta}_k$ to the explanatory variables and random effects, the corresponding additive model is given by

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} \mathbf{Z}_{jk} \gamma_{jk} \,. \qquad (1)$$

Here, $\boldsymbol{\theta}_k$ and $\boldsymbol{\eta}_k$ are vectors of length n, $\mathbf{X}_k$ is a known design matrix, $\boldsymbol{\beta}_k = (\beta_{1k}, \ldots, \beta_{J'_k k})$ is a parameter vector of length J'_k, $\mathbf{Z}_{jk}$ is a known design matrix and γ_{ik} is q_{ik}-dimensional random variable.

In order to include semiparametric nonlinear effects in equation (1), set $\mathbf{Z}_{jk} = \mathbf{I}_n$, where $\mathbf{I}_n$ is an $n \times n$ identity matrix and let $\gamma_{ik} = \mathbf{h}_{jk} = h_{jk}(\mathbf{x}_{jk})$, where $\mathbf{h}_{jk}$ is the vector which evaluates an unknown function h_{jk} at $\mathbf{x}_{jk}$, where $\mathbf{x}_{jk}$ is a vector of length n. Then, h_{jk} can be e.g. approximated using smoothing splines in the estimation. Thus, the GAMLSS can incorporate parametric, semiparametric and random-effect terms.

For areal data, De Bastiani et al. (2018a) show how random effects can be expressed in a way to account for structured spatial effects. The basis Gaussian Markov Random Fields (GMRF). Generally, a neighbourhood structure can be given by an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, that consists of vertices $\mathcal{V} = (1,2,\ldots,q)$ and a set of edges $\mathcal{E}$. A typical edge of the graph is (m,t), $t, m \in \mathcal{V}$. With respect to the graph, a random vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^T$ is called a GRMF with mean $\boldsymbol{\mu}$ and symmetric precision matrix $\lambda\mathbf{G}$, if and only if its density is given by

$$\pi(\boldsymbol{\gamma}) \propto exp\left[-\frac{1}{2}\lambda(\boldsymbol{\gamma} - \boldsymbol{\mu})^T \mathbf{G}(\boldsymbol{\gamma} - \boldsymbol{\mu})\right] \qquad (2)$$

and

$$G_{mt} \neq 0 \Leftrightarrow (m,t) \in \mathcal{E} \quad \text{for} \quad m \neq t, \qquad (3)$$

where G_{mt} is the element of matrix $\mathbf{G}$ for row m and column t (Rue and Held 2005). $\mathbf{G}$ contains the information about adjacent regions. When $\mathbf{G}$ is a non-singular matrix, the GMRF model is called conditional autoregressive (CAR) model (Besag 1974) and can be defined by

$$\gamma_i|\boldsymbol{\gamma}_{-i} \sim N\left(\sum_j \alpha_{ij}\gamma_j, k_i\right), \qquad (4)$$

where $\gamma_{-i} = (\gamma_1, \gamma_2, \dots, \gamma_{i-1}, \gamma_{i+1}, \dots, \gamma_q)$, $\alpha_{ii} = 0$, $\alpha_{ij} = -G_{ij}/G_{ii}(i \neq j)$ and $k_i = 1/(\lambda G_{ii})$, for $i = 1,2, \dots, q$. If $\mathbf{G}$ is symmetric, then $\boldsymbol{\mu} = \mathbf{0}$. For GAMs, typically a limiting case of the CAR, the intrinsic autoregressive model (IAR) is used to model spatially structured random effects (De Bastiani et al. 2018b). In order to incorporate an IAR model in the GAMLSS, its respective $\mathbf{Z}$ is set to be an index matrix indicating which observation belongs to which region. Then $\boldsymbol{\gamma}$ is a vector of q spatial random effects and $\boldsymbol{\gamma} \sim N(0, \lambda^{-1}\mathbf{G}^{-1})$. The intuitive interpretation is that such an effect follows Tobler's law and that fitted values from neighbouring regions are closer together. For more details see De Bastiani et al. (2018a).

The inferential framework for the estimation of a GAMLSS is derived from an empirical Bayesian argument. Assuming independent normal priors for γ_{jk}, it can be shown that the maximum-a-posteriori (MAP) is equivalent to penalized likelihood estimation for fixed smoothing (or hyper-) parameters. For this, algorithms relying on backfitting methods are used. These can be nested into methods for the estimation of the hyperparameters, which allows for an automated determination of the model's smoothing parameters (Rigby and Stasinopoulos 2005).

The designated dependent variable (RPR) in the present study is logically restricted to the $(0,1)$-interval[1]. Here, the appropriate choice for the variable's distribution is the Beta-distribution. The Beta-distribution is defined by two parameters and allows for a lot of flexibility (see Rigby et al. 2019). One way to parameterize the probability density function of the Beta-distribution is:

$$f_Y(y|\mu, \sigma) = \frac{1}{B(\alpha, \beta)} y^{\alpha-1}(1-y)^{\beta-1}. \qquad (5)$$

The parameters μ and σ are the location and scale parameter. They refer to the mean and the standard deviation of the variable Y. In this parameterization, $\alpha = \mu(1-\sigma^2)/\sigma^2$ and $\beta = (1-\mu)(1-\sigma^2)\sigma^2$, while $0 < \mu < 1$ and $0 < \sigma < 1$. $B(\alpha, \beta)$ represents the Beta-function. The mean of Y is given by $E(Y) = \mu$, the variance by $Var(Y) = \sigma^2\mu(1-\mu)$ (Rigby et al., 2019). Thus, in the present study two parameter vectors, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are estimated. The respective predictors are $\boldsymbol{\eta}_\mu$ and $\boldsymbol{\eta}_\sigma$.

[1] Under the plausible assumption that the rental rate will always be smaller than the sale price.

3 Data description and variable selection

3.1 Data description and processing

In order to calculate the RPR for a given plot of land, information both on its rent and sales price would be the best data base. As each plot is usually either sold or rented out at a given point in time, such information is usually not available. Therefore, local averages of land rents and prices are used as proxies. Therefore a dataset compiled from two sources was used. This first source is the data of the German agricultural census 2010 (*Landwirtschaftszählung* 2010). The second one is a dataset provided by the OGA Lower Saxony. While the first dataset is the most comprehensive dataset on the agricultural structure in Germany and is the latest available full survey. It contains information on all farms, including data on the land rents paid by the farms. The second dataset consist of plot level data of all agricultural land sales in Lower Saxony during the time period corresponding to the reference period of the agricultural census (second quarter of 2008 until first quarter of 2010). This data was originally collected by the OGA for the purposes of the German Federal Building Code.

The local average land rents, respectively land prices per ha were calculated on a standardised spatial grid which is the smallest grid used for official agricultural statistical purposes in Germany (e.g. Destatis n.d.). This grid has a cell size of 5x5 km and was used to merge the two data sources. Then the RPR of the respective cell was calculated. Noteworthy, within Germany, there is the major distinction between “arable land” and “grassland” (or “pasture”). The difference is that grassland is considered to be permanently used for forage production and that there are legal restrictions of its ploughing. Still, this does not imply that arable land is not used for forage production at all. Nevertheless, the profitability of the investment in land may vary between the two types. As the data allowed for a differentiation between arable and grassland on the cell level, the RPR was calculated separately. This lead to a total 2,702 local observations of the RPR. In order to allow for a differentiation between the land type of each observation of the RPR a cell level dummy variable *d_grassl* (= 1 if the observation is based on grassland land, 0 otherwise) was included in the final dataset. All further considered explanatory variables were calculated on the district level, using the farm level data of the agricultural census. The variables considered for the variable selection procedure was motivated by previous research on the determinants of rental rates for land (Habermann and Ernst 2010; Habermann and Breustedt 2011; März et al. 2016).

Overall, the considered variables reflected two key dimensions: (a) the district level agricultural structure and (b) the average production program in the district. With respect to (a), the share of farms, which are legal entities (*farm_type_share*), the share of part time farms (*parttime_share*), the share of rented land on the agricultural land (*rent_share1*), the average share of rented land on agricultural land per farm (*rent_share2*) and the share of organic farms (*organic_share*) in the district were calculated. Further, the labour intensity per ha (*labour*) and the average farm size (*size*) were considered. In order to account for the district level competition, a concentration measure was included in the analysis. It was defined in the form of a Herfindahl-Hirschmann-Index (*hhi*), and calculated as

$$hhi_j = \sum_{i=1}^{N_j} \left(\frac{Area\_of\_farm_i}{Total\_area_j} \right)^2 , \qquad (6)$$

for the *N* farms in a district *j* and measured for the concentration of farmland in that district. If each farm in the district would have the same size, the index would be equal to $\frac{1}{N}$ (thus going towards 0 with an increasing number equal sized farms) and would be equal to 1 if there is only one farm in the district (thus the land is fully concentrated).

With respect to (b), the average density of cattle (*cattle*), as well as hogs and poultry (*hog_poultry*) per ha were calculated (in animal units (AU)). In terms of crop production, shares of *potato*, *rye*, *sugar beet* and *winter wheat* in the cropping pattern were considered for the analysis. As discussed above, the potential effects of biogas production on land markets have gained interest in recent years (Habermann and Breustedt 2011; Hennig and Latacz-Lohmann 2017). To account for such potential effects, two variables reflecting biogas production in terms of the agricultural production direction as well as the farming structure were considered for the analysis. For the former case this is the district's average biogas capacity (in kWh) per ha (*biogas_cap*), for the latter it is the share of farms with biogas plants in the district (*biogas_share*). Lastly, the share of pasture on the total agricultural land in the district (*grass_share*) was considered. This variable is linked to the district's average production system. The considered variables are summarised in Table 1. All variables were considered both for the predictor for the mean and the scale parameter of the RPR. As discussed earlier, potential remaining spatial heterogeneity could be addressed by including structured spatial effects (f_{str}) and unstructured spatial effects (f_{unstr}) in the predictors.

Table 1: Variables considered for the analysis

Variable	Description
biogas_cap	average biogas capacity (in kWh) per ha
biogas_share	share of farms with biogas plants
cattle	average cattle density in animal units (AU) per ha
d_grassl	dummy variable, 1 if the observation refers to grassland, 0 otherwise
farm_type_share	share of legal entity farms
grassl_share	share of grassland on total agricultural land in production
hhi	Herfindahl-Hirschman index, based on the farm size in ha and the total amount of land under production in the district
hog_poultry	average density of hogs and poultry in AU per ha
labour	average labour force per ha
organic_share	share of organic farms
parttime_share	share of part-time farms
potato	share of potato in the cropping pattern
rent_share1	share of rented land on total agricultural land
rent_share2	average share of rented land on agricultural land per farm
rye	share of rye in the cropping pattern
size	average farm size in ha
sugarbeet	share of sugar beet in the cropping pattern
winterwheat	share of winter wheat in the cropping pattern

3.2 Variable selection

One of the major advantages of GAMLSS, is the ease to study complex models with different effect types. This also represents a potential drawback, as overly complex models are prone to overfitting, leading to an inadequate model. One way to select the terms to be included in the model are model comparisons based on the generalised Akaike information criterion (GAIC) (Stasinopoulos et al. 2017). Therefore, a modification of the procedure outlined by De Bastiani et al. (2018a) was applied. In a first step, an appropriate set of variables was selected. In a second step it was evaluated whether remaining heterogeneity could be explained by spatial effects. The procedure was as follows:

1. Estimate a 'Null model', containing only a constant in $\boldsymbol{\eta}_\mu$ and $\boldsymbol{\eta}_\sigma$.
2. Select variables to be included in the model, based on the GAIC.
   2.1. Apply a forward-stepwise-variable selection-procedure on $\boldsymbol{\eta}_\mu$.
   2.2. Apply a forward-stepwise-variable-selection-procedure on $\boldsymbol{\eta}_\sigma$, given the model obtained by step 2.1.
   2.3. Apply a backward-variable-elimination-procedure on the variables in $\boldsymbol{\eta}_\mu$, given model obtained by step 2.2.
3. Use the model obtained by step 2.3. and re-estimate the model, including
   3.1. an structured spatial effect in the predictor for $\boldsymbol{\eta}_\mu$,
   3.2. an structured spatial effect in the predictor for $\boldsymbol{\eta}_\sigma$,
   3.3. an structured spatial effect in the predictors for both $\boldsymbol{\eta}_\mu$ and $\boldsymbol{\eta}_\sigma$ and
   3.4. select the most appropriate model based on the GAIC.
4. Use the model obtained by step 3.4. and re-estimate the model, including
   4.1. an unstructured spatial effect in the predictor for $\boldsymbol{\eta}_\mu$,
   4.2. an unstructured spatial effect in the predictor for $\boldsymbol{\eta}_\sigma$,
   4.3. an unstructured spatial effect in the predictors for both $\boldsymbol{\eta}_\mu$ and $\boldsymbol{\eta}_\sigma$ and
   4.4. select the most appropriate model based on the GAIC.

The model obtained in step 4.4. will be used as the final model for the analysis. In case that the variable *d_grassl* is selected in Step 1, it is reasonable to also control for potential interaction effects between the variable *d_grassl* and other variables included in the model. These potential interactions are considered by the selection algorithm applied in the Steps 2.1 to 2.3. The final model selected by the procedure is presented and discussed in the following section.

4 Analysis of the rent-price-ratio

4.1 Results of the model selection and overview of the studied model

In this subsection, first the results of the model selection procedure are shown. The selected variables and interaction effects are presented in Table 2. Estimations were done using the 'R'-

software-package[2] (R Core Team 2019). It is important to note that regular standard errors obtained by the GAMLSS implementation may not be accurate when the model includes additive smoothing terms. Additionally, the standard errors do not account for the variable selection procedure, which further renders the interpretation of these effects unreliable (Hastie et al. 2009). In order to be able to assess the statistical significance, an additional non-parametric bootstrap procedure was carried out (Stasinopoulos et al. 2017).

Table 2: Variables selected in the GAMLSS-model (N=2,702)

Variable	μ	σ
biogas_share		
biogas_cap		√
cattle		
d_grassl	√	√
farm_type_share		√
grass_share		√
hhi	√	
parttime_share		
hog_poultry	√	
labour		√
organic_share	√	√
potato		√
rent_share1		
rent_share2		
rye		√
size	√	√
sugarbeet	√	
winterwheat		√
f_{str}	√	√
f_{unstr}	√	√
Interaction effects[a]		
d_grassl $*$ *hog_poultry*	√	
d_grassl $*$ *size*		√
d_grassl $*$ *labour*		√
d_grassl $*$ *farm_type_share*		√
d_grassl $*$ *winterwheat*		√
d_grassl $*$ *grass_share*		√
d_grassl $*$ *organic_share*		√
d_grassl $*$ *biogas_cap*		√

Note: [a] all other potential interaction effects were not selected in the model.

2 For the estimations of the models, the 'gamlss'-package (Rigby and Stasinopoulos 2005; Stasinopoulos and Rigby 2007) was used. The structured spatial effects rely on the implementation of the 'gamlss.spatial'-package (De Bastiani et al. 2018a). The bootstrap procedure was implemented using functions of the 'boot'-package (Davison and Hinkley 1997; Canty and Ripley 2019).

The selected variables represent both the local farming structure (e.g. *farm_type_share* and *labour*) and the production program (e.g. *sugarbeet* and *potato*). In particular, *d_grassl* was selected for the mean and scale predictor. In total, more variables were selected for the scale predictor then for the mean predictor. Further, a series of interaction effects between *d_grassl* and other variables were selected for both predictors. These effects are discussed in subsection 4.2. In both predictors, structured and unstructured spatial effects were included in the model. This is an indication that, after controlling for the selected variables, remaining heterogeneity has spatial components. These effects are discussed in subsection 4.3. The final model uses 55.94 degrees of freedom for the fit and yields a Global Deviance of -18,761.82, with a GAIC of -18,649.

4.2 Variable effects

The effects of the selected variable on both the mean and the scale of the RPRs distribution are presented and discussed. For the interpretation of the variable effects, it is helpful to recall that the RPR is a crude measure for the profitability of an investment. Thus, taking the perspective of a potential investor (and less of a farmer) as a potential buyer, allows for an intuitive interpretation of the results. Also, while the analysis shares some similarities with hedonic pricing studies, the results are not directly comparable, as the present study researched a relative and not an absolute measure. Table 3 shows the effects of the variables selected in $\boldsymbol{\eta}_{\mu}$ and thus, the effects on the average profitability of an investment in land (on the district level). Additionally, the respective 95 % confidence intervals (CI) presented, which are used to assess the statistical significance of the individual effects.

Table 3: Variable effects on the mean parameter (N=2,702)

Variable	β	lower 95 % CI	upper 95 % CI
Intercept	-4.3565	-4.6126	-4.2763
d_grassl	0.0994	0.0269	0.1358
hhi	-18.9214	-29.4627	-9.8509
hog_poultry	-0.0319	-0.0859	0.0876
organic_share	2.1994	-0.8389	3.9084
size	0.0078	0.0058	0.0119
sugarbeet	-1.7792	-2.2506	-1.1534
Interaction effects			
d_grassl $*$ *hog_poultry*	-0.2749	-0.3558	-0.1715

Notes: CI: Confidence interval, based on 100 bootstrap samples; Sources: RDC of the Federal Statistical Office and Statistical Offices of the Länder, agricultural census 2010 and data from the OGA Lower Saxony, own calculations.

The effect of the *hhi* was found to be significantly negative. Given its effect size, and the construction of the underlying variable (the higher the value, the higher the concentration of agricultural land), this implies that the average profitability of an investment in land was lower in districts with a more concentrated farming structure. At the same time, the effect of *size* indicates that the profitability of land investments was higher in districts with larger farms. This implies, that with respect to farming structure, there was a trade-off between average farm size (where economies of scale can increase returns) and land concentration (lowering competition and potentially inducing local market power). While the share of organic farms (*organic_share*) and the hog and poultry density (*hog_poultry*) were selected in the model, their CIs are too wide to be considered of significance. Interestingly, the results show that in districts with higher shares of *sugarbeet* in the cropping pattern, the average profitability of land investments was lower. This cannot be attributed to the returns of sugar beet productions (which are usually among the highest realizable in German agriculture, should increase the RPR). One explanation could be that these high returns influence land buyers' expectations of the returns and thus driving up land prices. The results show that the average profitability in investments in land is higher for grassland (*d_grassl*). Interestingly, there is a statistically significant interaction effect between *d_grassl* and *hog_poultry*, which indicates that investments in pasture were less profitable in animal refinement regions.

Table 4: Variable effects on the scale (N=2,702)

Variable	β	lower 95 % CI	upper 95 % CI
Intercept	0.9036	-6.2738	4.2629
biogas_cap	-0.3258	-1.0901	0.2833
d_grassl	-2.7099	-8.4173	6.6097
farm_type_share	-3.3419	-6.6835	3.9839
grassl_share	0.3654	-0.0343	1.2180
labour	-0.6323	-1.1764	-0.0011
organic_share	4.9366	1.1882	8.2266
potato	4.6136	2.6327	8.7017
rye	-1.8318	-4.3387	-0.6076
size	0.0048	-0.0023	0.0122
winterwheat	-1.1127	-2.1204	-0.2646
Interaction effects			
d_grassl ∗ *biogas_cap*	-0.8876	-2.1122	0.3626
d_grassl ∗ *farm_type_share*	3.8628	-5.2951	9.0133
d_grassl ∗ *grass_share*	-0.3803	-1.3736	0.0892
d_grassl ∗ *labour*	-0.3915	-1.2168	0.3314
d_grassl ∗ *organic_share*	-5.3019	-9.6347	0.3795
d_grassl ∗ *size*	0.0060	-0.0046	0.0131
d_grassl ∗ *winterwheat*	0.2822	-0.9928	1.3373

Notes: CI: Confidence interval, based on 100 bootstrap samples; Sources: RDC of the Federal Statistical Office and Statistical Offices of the Länder, agricultural census 2010 and data from the OGA Lower Saxony, own calculations.

The heterogeneity of the RPR (cf. Table 4) increased with the share of organic farms in the district (*organic_share*). Also, the results indicate that the RPRs heterogeneity increased with the share of *potato* production in the district. As potatoes, on the one hand, generate high returns, but on the other hand require relatively specific conditions. Hence the results appeare reasonable. The negative effects of *rye* and *winterwheat* could be interpreted in an analogue manner. As both crops require less specific production conditions, a higher share in the average production program can be seen as an indicator for more homogenous natural conditions. This would also indicate more homogenous returns in the district. The negative effect of *labour* could readily be explained by the fact that labour is a costly production factor which has to be compensated.

The remaining included variables which had wide CIs covering zero, thus it cannot be stated that they significantly contribute to the explanation of the profitability heterogeneity. Likewise, this can be seen as an indication that the local profitability heterogeneity did not significantly vary between pasture and arable land (*d_grassl*). While there are multiple interaction effects between the dummy variable indicating pasture land, and other variables (mostly referring to the farming structure), their CIs indicate that they were not statistically significant.

Lastly, it is interesting to reconsider the variables not selected in the model. Under the assumption that the RPR indicates the profitability of an investment, it appears reasonable, that the variables referencing the share of rented land and the share of part time farmers in the district were not selected in the model. Both biogas-production-related variables were not included in the model or cannot be considered being statistically significant. As previous research has indicated, rental rates were increased by biogas production (Hennig et al. 2014; Hennig and Latacz-Lohmann 2017). In context of the results of this paper, this suggests land prices must have increased by the same degree and indicates that biogas production did not influence market efficiency. This interpretation also holds for other variables not selected in the model.

While these results are readily interpreted from an investor's perspective, an interpretation from the farmer's perspective is more challenging. This is because the ratio of rental rates and sales prices does not contain any information about the absolute levels of prices and rents. Still, the results could be used to give some decision support. As farmers would usually either buy or rent land, the RPR can serve as an indicator whether it is relatively 'cheaper' to buy or rent land. Still, the final investment decision would still have to be made individually, as the actual profitability of an investment would depend on the realisable returns from farming activities.

It also has to be taken into account that the results (analogously to hedonic price studies) can include expectations of actors on the land market. Nevertheless it could be argued that if such expectations were present, they will most likely be expectations of the farming sector and not non-agricultural investors. This can be explained as the results covered a period prior to increased activities of such actors. Related herewith, the data on land rents averages information on all rental contracts, including potential older, long-running contracts. This is a fundamental issue of data availability, as land sales and rental contracts are commonly not observed at the same point in time[3].

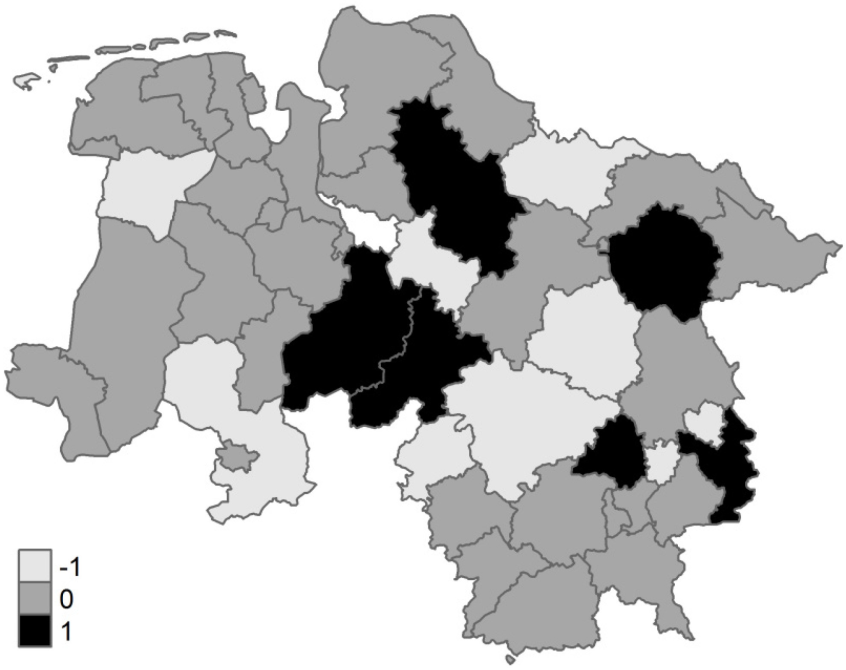
4.3 Spatial effects

As shown in Table 2, both structured and unstructured spatial effects were selected in the final model. In order to assess the significance of the spatial effects on the level of the individual district, CIs were again obtained using a nonparametric bootstrap-procedure. Generally, statistically significant district level effects were found for all four spatial effects terms. Nevertheless, when looking at the absolute size of the effects, it was found that the structured spatial effects were up to six orders of magnitude smaller when comparing the structured and unstructured effects for a given district. This leads to the conclusion, that in the chosen model

[3] The only hypothetical exception would be a case where a plot of land (without a current rental contract) is sold to an investor, who immediately rents it out to a farm.

selection procedure, the inclusion of both effects led to the best model (in terms of the GAIC). This happened whilst the spatial heterogeneity was mainly explained by the unstructured effects. Therefore, the structured spatial effect appeared to have no practical significance. The remainder of the subsection therefore focusses on the unstructured spatial effect.

Figure 1: Significance of the unstructured spatial effect for the district level mean of the rent-price-ratio for agricultural land based on bootstrapped confidence intervals

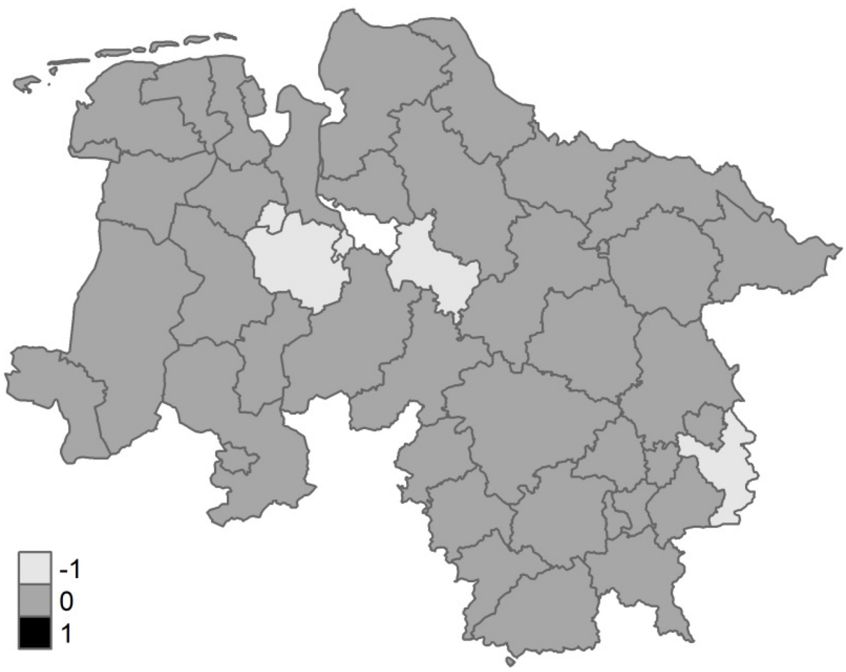


Note: -1 significantly negative, 0: non-significant, 1: significantly positive; the empty polygon represents the federal state Bremen; Sources: RDC of the Federal Statistical Office and Statistical Offices of the Länder, agricultural census 2010 and data from the OGA Lower Saxony, own calculations.

The unstructured district level effects on the mean and scale predictor are depicted in Figure 1 and Figure 2, respectively. In both Figures light grey indicates effects with CIs below zero, while black indicate CIs above zero. Grey indicates districts where the effects' CI overlap zero. Ceteris paribus, this can be interpreted as such that the respective parameter values of the RPR's distribution in a given district are larger or smaller. Generally, a higher number of statistically significant district level effects was found for the mean predictor (Figure 1). While they in principle represent unstructured remaining spatial heterogeneity, it became apparent, that most of the significant effects where in proximate to major urban centres (the neighbouring federal (city) states Bremen and Hamburg, as well as the capital of the federal state, Hannover). Nevertheless these effects were still unstructured, as no clear pattern with respect to the effect direction emerged. The majority of the district level unstructured effects for the scale predictor had been considered not statistically significantly different from zero (Figure 2). All of the exceptions have be considered to be negative effects. Again, these unstructured effects reflect local particularites. Here, most of them are close/around the city state Bremen.

Summarising, these results show that there are likely some general influences from urban centers, while still their consequences remain district specific.

Figure 2: Significance of the unstructured spatial effect for the district level scale of the rent-price-ratio for agricultural land based on bootstrapped confidence intervals



Note: -1 significantly negative, 0: non-significant, 1: significantly positive; the empty polygon represents the federal state Bremen; Sources: RDC of the Federal Statistical Office and Statistical Offices of the Länder, agricultural census 2010 and data from the OGA Lower Saxony, own calculations.

5 Conclusion

By focussing on the cross-sectional profitability of investments in farm land, the results presented in this paper complement and extend prior research. The district level analysis of the RPR of agricultural land revealed that both the average RPR as well as its heterogeneity were influenced by the districts average production program and the farming structure. Additionally, remaining district level spatial heterogeneity was identified.

The results have some implications for policy makers, farmers and other actors on land markets. Overall, investments in districts with a higher land concentration were less profitable. One potential interpretation is that the local concentration may reached levels at which competition on the land markets reduced to a level at which individual farms gain market power. In light of political and societal discussions about the effects of structural change and land ownership in the agricultural sector, this would be particularly relevant. Still, present results could not be used to fully justify this interpretation. Here, more research is needed. Furthermore, the results have some implications for future research.

In future research, other model specifications, e.g. with nonlinear effects, could be considered. Also, other variable selection mechanisms (like boosting) could be considered. Given the recent developments of the land markets, further insights could be gained by expanding the dataset, either in the spatial, temporal or even spatiotemporal domain, especially as corresponding additive terms could directly be incorporated in the model. When more recent data (e.g. from the next agricultural census) becomes available, it would be particularly interesting to investigate whether the activity of non-agricultural investors altered the relationships found in the present study.

6 References

Alston, J.M. 1986. “An Analysis of Growth of U.S. Farmland Prices, 1963–82.” *American Journal of Agricultural Economics* 68(1):1–9. doi: 10.2307/1241644.

Besag, J. 1974. “Spatial Interaction and the Statistical Analysis of Lattice Systems.” *Journal of the Royal Statistical Society: Series B (Methodological)* 36(2):192–225. doi: 10.1111/j.2517-6161.1974.tb00999.x.

Burt, O.R. 1986. “Econometric Modeling of the Capitalization Formula for Farmland Prices.” *American Journal of Agricultural Economics* 68(1):10–26. doi: 10.2307/1241645.

Campbell, J.Y., and R.J. Shiller. 1988. “The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors.” *The Review of Financial Studies* 1(3):195–228. doi: 10.1093/rfs/1.3.195.

Canty, A., and B.D. Ripley. 2019. *Boot: bootstrap R (s-plus) functions*.

Clark, J.S., M. Fulton, and J.T. Scott. 1993. “The Inconsistency of Land Values, Land Rents, and Capitalization Formulas.” *American Journal of Agricultural Economics* 75(1):147–155. doi: 10.2307/1242963.

Davison, A.C., and D.V. Hinkley. 1997. *Bootstrap methods and their applications*. Cambridge: Cambridge University Press.

De Bastiani, F., R.A. Rigby, D.M. Stasinopoulous, A.H.M.A. Cysneiros, and M.A. Uribe-Opazo. 2018. “Gaussian Markov random field spatial models in GAMLSS.” *Journal of Applied Statistics* 45(1):168–186. doi: 10.1080/02664763.2016.1269728.

De Bastiani, F., M. Stasinopoulos, and R. Rigby. 2018. *Gamlss.spatial: spatial terms in generalized additive models for location scale and shape models*. URL: https://CRAN.R-project.org/package=gamlss.spatial.

Destatis n.d. “Atlas Agrarstatistik.” URL: https://www.atlas-agrarstatistik.nrw.de/ [Accessed: 12 December 2019].

Eurostat. 2019. “Agricultural land prices by region.” URL: https://ec.europa.eu/eurostat/web/products-datasets/product?code=apri_lprc [Accessed: 11 December 2019].

Fahrmeir, L., and T. Kneib. 2011. *Bayesian Smoothing and Regression for Longitudinal, Spatial and Event History Data*. Oxford, New York: Oxford University Press.

Falk, B. 1991. “Formally Testing the Present Value Model of Farmland Prices.” *American Journal of Agricultural Economics* 73(1):1–10. doi: 10.2307/1242877.

Feichtinger, P., and K. Salhofer. 2013. “What do we know about the influence of agricultural support on agricultural land prices?” *German Journal of Agricultural Economics* 62(2):1–15.

Gutierrez, L., J. Westerlund, and K. Erickson. 2007. “Farmland prices, structural breaks and panel data.” *European Review of Agricultural Economics* 34(2):161–179. doi: 10.1093/erae/jbm018.

Habermann, H., and G. Breustedt. 2011. “Einfluss der Biogaserzeugung auf landwirtschaftliche Pachtpreise in Deutschland.” *German Journal of Agricultural Economics* 51(2):85–100.

Habermann, H., and C. Ernst. 2010. “Entwicklungen und Bestimmungsgründe der Landpachtpreise in Deutschland.” *Berichte über Landwirtschaft* 88(1):57–85.

Hallam, D., F. Machado, and G. Rapsomanikis. 1992. “Co-Integration Analysis and the Determinants of Land Prices.” *Journal of Agricultural Economics* 43(1):28–37. doi: 10.1111/j.1477-9552.1992.tb00195.x.
Hastie, T., and R. Tibshirani. 1990. *Generalized additive models*. Boca Raton, Fla: Chapman & Hall/CRC.
Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition* 2nd ed. New York: Springer-Verlag.
Hennig, S., G. Breustedt, and U. Latacz-Lohmann. 2014. “The impact of payment entitlements on arable land prices and rental rates in Schleswig-Holstein.” *German Journal of Agricultural Economics* 63(4):219–239.
Hennig, S., and U. Latacz-Lohmann. 2017. “The incidence of biogas feed-in tariffs on farmland rental rates – evidence from northern Germany.” *European Review of Agricultural Economics* 44(2):231–254. doi: 10.1093/erae/jbw023.
Hüttel, S., L. Wildermann, and C. Croonenbroeck. 2016. “How do institutional market players matter in farmland pricing?” *Land Use Policy* 59:154–167. doi: 10.1016/j.landusepol.2016.08.021.
Hyder, K., and A.H. Maunder. 1974. “The price of farms.” *Oxford Agrarian Studies* 3(1):3–14. doi: 10.1080/13600817408423812.
Ibendahl, G., and T. Griffin. 2013. “The Connection Between Cash Rents and Land Values.” *Journal of the American Society of Farm Managers and Rural Appraisers (ASFMRA)*:239–247.
Latruffe, L., and C.L. Mouël. 2009. “Capitalization of Government Support in Agricultural Land Prices: What Do We Know?” *Journal of Economic Surveys* 23(4):659–691. doi: 10.1111/j.1467-6419.2009.00575.x.
März, A., N. Klein, T. Kneib, and O. Musshoff. 2016. “Analysing farmland rental rates using Bayesian geoadditive quantile regression.” *European Review of Agricultural Economics* 43(4):663–698. doi: 10.1093/erae/jbv028.
Mishra, A.K., and C.B. Moss. 2013. “Modeling the effect of off-farm income on farmland values: A quantile regression approach.” *Economic Modelling* 32:361–368. doi: 10.1016/j.econmod.2013.02.022.
NMELV (Niedersächsisches Ministerium für Ernährung, Landwirtschaft und Verbraucherschutz). 2017. “Die niedersächsische Landwirtschaft in Zahlen 2017.” URL: https://www.ml.niedersachsen.de/download/124920/Die_niedersaeschsische_Landwirtschaft_in_Zahlen_2017.pdf.
Phipps, T.T. 1984. “Land Prices and Farm-Based Returns.” *American Journal of Agricultural Economics* 66(4):422–429. doi: 10.2307/1240920.
Plogmann, J.M.; Mußhoff, O.; Odening, M.; Ritter, M. 2020." What Moves the German Land Market? A Decomposition of the Land Rent-Price Ratio". *German Journal of Agricultural Economics* (in press).
R Core Team. 2019. *R: A language and environment for statistical computing*. Vienna, Austria. URL: https://www.R-project.org/.
Rigby, R.A., and D.M. Stasinopoulos. 2005. “Generalized additive models for location, scale and shape (with discussion).” *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54(3):507–554. doi: 10.1111/j.1467-9876.2005.00510.x.
Rigby, R.A., M.D. Stasinopoulos, G.Z. Heller, and F. De Bastiani. 2019. *Distributions for Modelling Location, Scale, and Shape Using GAMLSS in R.* Milton: CRC Press LLC.
Rue, H., and L. Held. 2005. *Gaussian Markov random fields: theory and applications*. Boca Raton: Chapman & Hall/CRC.
Saguatti, A., K. Erickson, and L. Gutierrez. 2014. “Spatial panel models for the analysis of land prices.” Paper presented at the EAAE 2014 Congress in Ljubljana, Slovenia. doi: 10.22004/ag.econ.172997.
Stasinopoulos, D.M., and R.A. Rigby. 2007. “Generalized additive models for location scale and shape (GAMLSS) in R.” *Journal of Statistical Software* 23(7):1–46.
Stasinopoulos, M.D., G.Z. Heller, V. Voudouris, and F. De Bastiani. 2017. *Flexible regression and smoothing: using GAMLSS in R*. Boca Raton: CRC Press/Taylor & Francis Group.

Statistisches Bundesamt. 2018. “Kaufwerte für landwirtschaftliche Grundstücke.” Fachserie 3 No. Reihe 2.4-2018. URL: https://www.destatis.de/DE/Themen/Wirtschaft/Preise/Baupreise-Immobilienpreisindex/Publikationen/Downloads-Bau-und-Immobilienpreisindex/kaufwerte-landwirtschaftliche-grundstuecke-2030240187004.pdf [Accessed: 11 December 2019].

Tietz, A., B. Forstner, and P. Weingarten. 2013. “Non-agricultural and supra-regional investors on the german agricultural land market: an empirical analysis of their significance and impacts.” *German Journal of Agricultural Economics* 62(2):1-13.

Traill, B. 1979. “An empirical model of the U.K. land market and the impact of price policy on land values and rents.” *European Review of Agricultural Economics* 6(2):209–232. doi: 10.1093/erae/6.2.209.

Turvey, C.G. 2003. "Hysteresis and the Value of Farmland: A Real-Options Approach to Farmland Valuation." In: Moss, C.B., Schmitz, A. (eds.): *Government Policy and Farmland Markets*, Iowa State Press: 179-207.

Umlauf, N., and T. Kneib. 2018. “A primer on Bayesian distributional regression.” *Statistical Modelling* 18(3–4):219–247. doi: 10.1177/1471082X18759140.

Yang, X., M. Ritter, and M. Odening. 2017. “Testing for regional convergence of agricultural land prices.” *Land Use Policy* 64:64–75. doi: 10.1016/j.landusepol.2017.02.030.